

Slide 1 of 15

Title Slide: Considerations for Analysis of NHES Data

Slide 2 of 15

This module describes the analytic considerations that should be kept in mind when using NHES data. It describes the similarities and differences of the NHES:2012 data files with past collections. This module also discusses specific considerations and data anomalies in the NHES:2012 that researchers should take into account when using the data files.

Slide 3 of 15

There are a few broad considerations that are important to keep in mind regarding the NHES data.

First, as mentioned previously, use caution when making comparisons between the 2012 data and prior years. The mode change from telephone to mail may confound your comparisons.

Second, the unit of analysis in the 2012 surveys is the child. All estimates describe the number and percentage of children, not parents or families.

Third, the NHES data, in all years, are cross-sectional. NHES data are not longitudinal. This means it is not possible to establish causality. The NHES provides a snapshot of the population in the data collection year.

Fourth, the NHES is a national survey. NHES data are not appropriate for state-level analysis. Sample sizes in the NHES will not support analysis at the state level. The variable that indicates the state of the sampled household is only available to users who hold a restricted-use data license. The lowest level of geographic information available on the public NHES data files is Census region.

Finally, in all years, the NHES was conducted in English and Spanish only.

The rest of this module describes the specific data considerations and anomalies in the NHES:2012 data files. All of these considerations are described in depth in the data file user's manual for 2012, which can be accessed by clicking on the corresponding underlined screen text.

Slide 4 of 15

When using the NHES data for trend analysis, there are several challenges that researchers should take into consideration.

The first is that similar items are often used in different Topical surveys and thus are asked of somewhat different populations. For example, the 2007 PFI asks about

Considerations for Analysis of NHES Data

school-age children but includes some School Readiness items for children in Kindergarten through second grade.

Another consideration is that pooling data across surveys is not advisable, due to the lack of appropriate weights. Analysts do not need to pool NHES data in order to test differences in estimates between years. NHES data in different years are treated as independent samples.

Finally, analysts using data from the NHES:2012 should note that the change in data collection mode resulted in changes in question wording for some items that were administered in the past.

Slide 5 of 15

There are school-level derived variables available for analysis on the NHES:2012 PFI data file. For children enrolled in school, parents were asked to report the child's school. The reported school was then linked to data about that school available in NCES's Common Core of Data and Private School Universe Survey. As a result, some data from these two external sources were appended to the PFI data file. A detailed list of appended variables can be found in the data file user's manual. A few of these variables are provided in the public-use data; however, the majority are only found on restricted-use data.

It is important to note that the most recent CCD and PSS source data files available at the time of the appending were those from the 2009-2010 school year. Therefore, some of the data may have changed between the 2009-2010 school year and the time of the NHES collection during the 2011-2012 school year.

Any data that were missing in the source files were not imputed.

Slide 6 of 15

Special editing procedures were performed on household composition data collected in the NHES. This includes the variable HHTOTAL, which is the total number of people living in the household. It also includes the individual relationship variables detailing how each household member relates to the sampled child.

First, values of HHTOTAL greater than 8 were top-coded; that is, they were capped at a maximum value of 8. This top-coding was used to protect the confidentiality of respondents.

In cases where HHTOTAL exceeded the sum of the individual composition variables, a new variable HHUNID which stands for unidentified household members – was set to the difference. In cases where HHTOTAL was less than the sum of the individual composition variables, HHTOTAL was set to equal the sum of these variables.

These calculations were performed for both the PFI and ECPP data files.

Slide 7 of 15

In some cases, a parent or child was reported as Hispanic but their race was not reported. This situation occurred for around a quarter of all cases. Rather than impute these cases, new variables were created on both the PFI and ECPP data files to denote when ethnicity was reported as Hispanic, but race was missing. This variable for child's race is CHISPRM. For the first and second parent, this variable is named P1 and P2HISPRM, respectively.

Slide 8 of 15

All parent/guardian age variables have been top-coded at age 90. Also, the age when a parent or guardian first became a parent to any child has been bottom-coded at age 12. These edits were put in place to protect respondent confidentiality.

Slide 9 of 15

The way the NHES counts the number of homeschoolers in the United States was affected by the change in mode from telephone to mail in the NHES:2012. Researchers interested in homeschooled students should refer to the data file user's manual for detailed information about how this change affects this population. The main difference is in how part-time homeschoolers are counted. Some students were reported to be homeschooled and enrolled in school on the PFI Enrolled survey and are included in the count of homeschoolers in 2012. However, the NHES:2012 PFI Enrolled questionnaire does not allow researchers to identify how many hours per week these students are enrolled in a school. The number of hours per week that homeschooled children spent in school has historically been a factor that is included when defining the NHES homeschool population. In order to compare the 2012 homeschooling counts to prior years, a special weight should be used. This information is provided in the data file user's manual, which can be accessed by clicking on the corresponding underlined screen text.

Slide 10 of 15

There are several cases where the detailed relationships of mothers and fathers to the sampled children are unusual. For example, the child was reported to have a birth mother and foster father at home. Data users interested in foster parent relationships should consider how to handle these cases in their analyses.

Slide 11 of 15

There are some cases in the PFI data file where the child's age and grade in school do not appear to match. These include, for example, a 12-year-old in 12th grade, a 17-year-old in first grade, a 16-year-old in 3rd grade, and children over age 8 in Kindergarten. In these cases, the questionnaire was examined to ensure the data reflected the respondent's answer and was not the result of a keying error. These cases were confirmed as actual respondent reports and therefore were left as reported. Again, analysts should consider how to handle these cases in their analyses.

Slide 12 of 15

On the PFI data file, there are some cases where the parent-reported data about the type of school the child attends differs from the appended school data. Parent-reported data for these cases were not changed. These anomalies could have been due to parent misreporting of the type of school his/her child attends, misidentification of the school by the parent, problems with the school type data from either the CCD or PSS, or other unknown factors.

Slide 13 of 15

There was an error in the Spanish translation of civil unions and domestic partnerships which affected reporting on the marital and partner status questions in the parent sections of all Spanish-version questionnaires. This error affected both data files and led to significantly higher reports of civil unions or domestic partnerships in the Spanish versions than what would reasonably be expected in the population and compared to the English versions.

This error was accounted for in data editing by recoding responses of civil and domestic partnership to “married” when the questionnaire was completed in Spanish and there were two parents or guardians in the household who were of the opposite sex. This occurred for 34 cases on the ECPP file and 42 cases on the PFI file.

Slide 14 of 15

On both the PFI and ECPP data files, values of child’s place of birth in the variable CPLCBRTH, that were imputed using hot-deck imputation are skewed toward a child being born in another country compared to reported values. However, the change in the overall percentage distribution is minimal, so the NHES-standard hot-deck procedure was retained for this variable. Analysts who want to examine child’s place of birth in their research may wish to evaluate the missing data using the imputation flag variable and consider other imputation methods if desired.

Slide 15 of 15

This module described the analytic considerations that should be kept in mind when using NHES data. It described the similarities and differences of the 2102 NHES data files with past collections, as well as specific considerations and data anomalies in the NHES:2012 that researchers should take into account when using the data files.

Additionally, important resources that have been provided throughout the module are summarized here along with the module’s objectives for your reference.

This concludes the NHES dataset training. You may now click the exit button to return to the landing page.